

862.C2175



D.T.  
#6 3-29-02  
Priority Papers  
PATENT APPLICATION

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

RECEIVED

JUL 18 2001

Technology Center 2600

In re Application of:

YASUO OKUTANI, ET AL.

Application No.: 09/818,581

Filed: March 28, 2001

For: SPEECH SYNTHESIS APPARATUS AND  
METHOD, AND STORAGE MEDIUM

Examiner: Not Assigned

Group Art Unit: 2644

July 2, 2001

RECEIVED

JUL 19 2001

Technology Center 2600

RECEIVED

JUL 05 2001

Technology Center 2600

Commissioner for Patents  
Washington, D.C. 20231

CLAIM TO PRIORITY

Sir:

Applicants hereby claim priority under the International Convention and all rights  
to which they are entitled under 35 U.S.C. § 119 based upon the following Japanese Priority

Application:

JAPAN

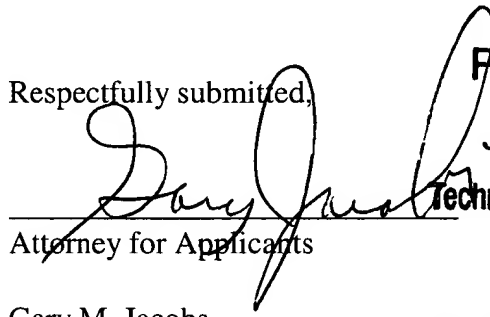
2000-099420

March 31, 2000.

A certified copy, with the translation of the first page of the priority document, is  
enclosed.

Applicants' undersigned attorney may be reached in our Washington, D.C. office by telephone at (202) 530-1010. All correspondence should continue to be directed to our address given below.

Respectfully submitted,

  
Attorney for Applicants

Gary M. Jacobs  
Registration No. 28,861

RECEIVED

JUL 1 8 2001

Technology Center 2600

RECEIVED

JUL 1 9 2001

Technology Center 2600

RECEIVED

JUL 0 5 2001

Technology Center 2600

FITZPATRICK, CELLA, HARPER & SCINTO  
30 Rockefeller Plaza  
New York, New York 10112-3801  
Facsimile: (212) 218-2200  
GMJ/cmv

(translation of the front page of the priority document of  
Japanese Patent Application No. 2000-099420)

PATENT OFFICE  
JAPANESE GOVERNMENT

This is to certify that the annexed is a true copy of the  
following application as filed with this Office.

Date of Application: March 31, 2000

Application Number : Patent Application 2000-099420

Applicant(s) : Canon Kabushiki Kaisha

April 20 2001

Commissioner,

Patent Office

Kouzo OIKAWA

Certification Number 2001-3033156

09/8/8,581  
YOSHINO OKUTANI, et al  
3-28-07

RECEIVED

JUL 18 2001

Technology Center 2600

RECEIVED

JUL 19 2001

Technology Center 2600

RECEIVED

JUL 05 2001

Technology Center 2600



日 本 国 特 許  
JAPAN PATENT OFFICE

09/818,581  
YOSHINO OKUTANI, et al.  
3-28-01

別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office

出 願 年 月 日

Date of Application:

2000年 3月31日

RECEIVED

JUL 18 2001

出 願 番 号

Application Number:

特願2000-099420

Technology Center 2600

出 願 人

Applicant(s):

キヤノン株式会社

RECEIVED

JUL 19 2001

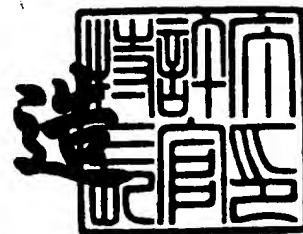
Technology Center 2600

CERTIFIED COPY OF  
PRIORITY DOCUMENT

2001年 4月20日

特 許 庁 長 官  
Commissioner,  
Japan Patent Office

及 川 耕 造



【書類名】 特許願

【整理番号】 4172011

【提出日】 平成12年 3月31日

【あて先】 特許庁長官殿

【国際特許分類】 G01L 5/04

【発明の名称】 音声情報処理装置とその方法と記憶媒体

【請求項の数】 21

【発明者】

    【住所又は居所】 東京都大田区下丸子3丁目30番2号 キヤノン株式会社  
社内

    【氏名】 奥谷 泰夫

【発明者】

    【住所又は居所】 東京都大田区下丸子3丁目30番2号 キヤノン株式会社  
社内

    【氏名】 小森 康弘

【特許出願人】

    【識別番号】 000001007

    【氏名又は名称】 キヤノン株式会社

【代理人】

    【識別番号】 100076428

    【弁理士】

    【氏名又は名称】 大塚 康徳

    【電話番号】 03-5276-3241

【選任した代理人】

    【識別番号】 100101306

    【弁理士】

    【氏名又は名称】 丸山 幸雄

    【電話番号】 03-5276-3241

【選任した代理人】

【識別番号】 100115071

【弁理士】

【氏名又は名称】 大塚 康弘

【電話番号】 03-5276-3241

【手数料の表示】

【予納台帳番号】 003458

【納付金額】 21,000円

【提出物件の目録】

【物件名】 明細書 1

【物件名】 図面 1

【物件名】 要約書 1

【包括委任状番号】 0001010

【プルーフの要否】 要

【書類名】 明細書

【発明の名称】 音声情報処理装置とその方法と記憶媒体

【特許請求の範囲】

【請求項 1】 音素素片を所定の韻律情報に基づいて編集することによって生じる歪を求める歪出力手段と、

音声合成に使用する素片辞書に登録する音声素片を、前記歪出力手段から出力された歪に基づいて選択する素片登録手段と、  
を有することを特徴とする音声情報処理装置。

【請求項 2】 前記歪出力手段は、前記音声素片を他の音声素片と接続することによって生じる接続歪と前記音声素片を変形することによって生じる変形歪とに基づいて、前記歪を求めることを特徴とする請求項 1 に記載の音声情報処理装置。

【請求項 3】 テキストデータを入力するテキスト入力手段と、  
前記入力されたテキストデータの言語解析を行なう言語解析手段と、  
前記言語解析手段による解析結果に基づいて前記所定の韻律情報を生成する韻律生成手段を更に有することを特徴とする請求項 1 又は 2 に記載の音声情報処理装置。

【請求項 4】 前記接続歪及び変形歪により決定される歪を基準として音声素片系列の Nbest 系列を求める Nbest 決定手段を更に有し、

前記素片登録手段は、前記音声素片系列の Nbest 系列を基に前記素片辞書に登録する音声素片を選択することを特徴とする請求項 2 又は 3 に記載の音声情報処理装置。

【請求項 5】 前記素片登録手段は、前記接続歪と前記変形歪との重み付き加算に基づいて、前記素片辞書に登録する音声素片を選択することを特徴とする請求項 2 又は 3 に記載の音声情報処理装置。

【請求項 6】 前記歪出力手段は、各音声素片のケプストラム距離を用いて前記接続歪を決定することを特徴とする請求項 2 乃至 5 のいずれか 1 項に記載の音声情報処理装置。

【請求項 7】 前記歪出力手段は、変形前の音声素片と変形後の音声素片に

おけるケプストラム距離を用いて前記変形歪を決定することを特徴とする請求項 2 乃至 5 のいずれか 1 項に記載の音声情報処理装置。

【請求項 8】 前記歪出力手段は、前記変形歪を記憶したテーブルを有し、当該テーブルを参照して前記変形歪を決定することを特徴とする請求項 2 乃至 5 のいずれか 1 項に記載の音声情報処理装置。

【請求項 9】 前記歪出力手段は、前記接続歪を記憶したテーブルを有し、当該テーブルを参照して前記接続歪を決定することを特徴とする請求項 2 乃至 5 のいずれか 1 項に記載の音声情報処理装置。

【請求項 1 0】 前記素片辞書を用いてテキストデータを音声合成する音声合成手段を更に有することを特徴とする 1 乃至 9 のいずれか 1 項に記載の音声情報処理装置。

【請求項 1 1】 音素素片を所定の韻律情報に基づいて編集することによって生じる歪を求める歪出力工程と、

音声合成に使用する素片辞書に登録する音声素片を、前記歪出力工程で出力された歪に基づいて選択する素片登録工程と、  
を有することを特徴とする音声情報処理方法。

【請求項 1 2】 前記歪出力工程は、前記音声素片を他の音声素片と接続することによって生じる接続歪と前記音声素片を変形することによって生じる変形歪とに基づいて、前記歪を求めることを特徴とする請求項 1 1 に記載の音声情報処理方法。

【請求項 1 3】 テキストデータを入力するテキスト入力工程と、  
前記入力されたテキストデータの言語解析を行なう言語解析工程と、  
前記言語解析工程による解析結果に基づいて前記所定の韻律情報を生成する韻律生成工程を更に有することを特徴とする請求項 1 1 又は 1 2 に記載の音声情報処理方法。

【請求項 1 4】 前記接続歪及び変形歪により決定される歪を基準として音声素片系列の Nbest 系列を求める Nbest 決定工程を更に有し、

前記素片登録工程では、前記 Nbest 系列を基に前記素片辞書に登録する音声素片を選択することを特徴とする請求項 1 2 又は 1 3 に記載の音声情報処理方法。



【請求項 1 5】 前記素片登録工程では、前記接続歪と前記変形歪との重み付き加算に基づいて、前記素片辞書に登録する音声素片を選択することを特徴とする請求項 1 2 又は 1 3 に記載の音声情報処理方法。

【請求項 1 6】 前記歪出力工程では、各音声素片のケプストラム距離を用いて前記接続歪を決定することを特徴とする請求項 1 2 乃至 1 5 のいずれか 1 項に記載の音声情報処理方法。

【請求項 1 7】 前記歪出力工程では、変形前の音声素片と変形後の音声素片におけるケプストラム距離として変形歪を定量化して決定することを特徴とする請求項 1 2 乃至 1 5 のいずれか 1 項に記載の音声情報処理方法。

【請求項 1 8】 前記歪出力工程では、前記変形歪を記憶したテーブルを有し、当該テーブルを参照して前記変形歪を決定することを特徴とする請求項 1 2 乃至 1 5 のいずれか 1 項に記載の音声情報処理方法。

【請求項 1 9】 前記歪出力工程では、前記接続歪を示すテーブルを有し、当該テーブルを参照して前記接続歪を決定することを特徴とする請求項 1 2 乃至 1 5 のいずれか 1 項に記載の音声情報処理方法。

【請求項 2 0】 前記素片辞書を用いてテキストデータを音声合成する音声合成工程を更に有することを特徴とする 1 1 乃至 1 9 のいずれか 1 項に記載の音声情報処理方法。

【請求項 2 1】 請求項 1 1 乃至 2 0 のいずれか 1 項に記載の方法を実行するプログラムを記憶したことを特徴とする、コンピュータにより読取り可能な記憶媒体。

【発明の詳細な説明】

【0 0 0 1】

【発明の属する技術分野】

本発明は、音声合成で使用される素片辞書を作成する音声情報処理装置及びその方法と記憶媒体に関するものである。

【0 0 0 2】

【従来の技術】

近年、音声素片を 1 ピッチ波形単位で複製及び、或いは削除しながら所望のピ

ッチ間隔で貼り合わせて編集し（PSOLA：ピッチ同期波形重畳法）、それらの音声素片を接続して音声合成する音声合成方法が主流となっている。

【0003】

【発明が解決しようとする課題】

このような技術を利用して音声合成された音声には、音声素片を編集することによる歪（以下、変形歪）と、音声素片を接続することによる歪（以下、接続歪）とが含まれる。これら2つの歪が、合成された音声の品質劣化を引き起こす大きな要因となる。中でも、素片辞書に登録できる音声素片の数が制限される状況下では、音声合成時に、このような歪が小さくなるように音声素片を選択する余地がほとんど残されていない場合がある。特に、一つの音韻環境について1つの音声素片しか素片辞書に登録できない場合には、歪が小さくなるように音声素片を選択する余地は全くなり、このような素片辞書を用いると、変形歪や接続歪による合成音声の品質劣化は避けられないものとなる。

【0004】

本発明は上記従来例に鑑みてなされたもので、接続歪や変形歪に基づき歪の影響を考慮して、素片辞書に登録する音声素片を選択することによって音声合成の音質劣化を抑制する音声情報処理装置及びその方法と記憶媒体を提供することを目的とする。

【0005】

【課題を解決するための手段】

上記目的を達成するために本発明の音声情報処理装置は以下のような構成を備える。即ち、

音素素片を所定の韻律情報に基づいて編集することによって生じる歪を求める歪出力手段と、

音声合成に使用する素片辞書に登録する音声素片を、前記歪出力手段から出力された歪に基づいて選択する素片登録手段と、  
を有することを特徴とする。

【0006】

また前記歪出力手段は、前記音声素片を他の音声素片と接続することによって

生じる接続歪と前記音声素片を変形することによって生じる変形歪とに基づいて、前記歪を求めることを特徴とする。

【0007】

上記目的を達成するために本発明の音声情報処理方法は以下のような工程を備える。即ち、

音素素片を所定の韻律情報に基づいて編集することによって生じる歪を求める歪出力工程と、

音声合成に使用する素片辞書に登録する音声素片を、前記歪出力工程で出力された歪に基づいて選択する素片登録工程と、  
を有することを特徴とする。

【0008】

【発明の実施の形態】

以下、添付図面を参照して本発明の好適な実施の形態を詳細に説明する。

【0009】

【実施の形態1】

図1は、本発明の実施の形態に係る音声情報処理装置のハードウェア構成を示すブロック図である。尚、本実施の形態では、一般的なパーソナルコンピュータを音声合成装置として用いる場合について説明するが、本発明は専用の音声情報処理装置であっても、また他の形態の装置であっても良い。

【0010】

図1において、101は制御メモリ（ROM）で、中央処理装置（CPU）102で使用される各種制御データを記憶している。CPU102は、RAM103に記憶された制御プログラムを実行して、この装置全体の動作を制御している。103はメモリ（RAM）で、CPU102による各種制御処理の実行時、ワークエリアとして使用されて各種データを一時的に保存するとともに、CPU102による各種処理の実行時、外部記憶装置104から制御プログラムをロードして記憶している。この外部記憶装置は、例えばハードディスク、CD-ROM等を含んでいる。105はD/A変換器で、音声信号を示すデジタルデータが入力されると、これをアナログ信号に変換してスピーカ109に出力して音声を再

生する。106は入力部で、オペレータにより操作される、例えばキーボードや、マウス等のポインティングデバイスを備えている。107は表示部で、例えばCRTや液晶等の表示器を有している。108はバスで、これら各部を接続している。110は音声合成ユニットである。

#### 【0011】

以上の構成において、本実施の形態の音声合成ユニット110を制御するための制御プログラムは外部記憶装置104からロードされてRAM103に記憶され、その制御プログラムで用いる各種データは、制御メモリ101に記憶されている。これらのデータは、中央処理装置102の制御の下にバス108を通じて適宜メモリ103に取り込まれ、中央処理装置102による制御処理で使用される。D/A変換器105は、制御プログラムを実行することによって作成される音声波形データ（デジタル信号）をアナログ信号に変換してスピーカ109に出力する。

#### 【0012】

図2は、本実施の形態に係る音声合成ユニット110のモジュール構成を示すブロック図で、この音声合成ユニット110は、大きく分けて、素片辞書206に音声素片を登録するための処理を実行する素片辞書作成モジュールと、テキストデータを入力し、そのテキストデータに対応する音声を合成して出力する処理を行なう音声合成モジュールの2つのモジュールを有している。

#### 【0013】

図2において、201は、入力部106又は外部記憶装置104から任意のテキストデータを入力するテキスト入力部、202は解析辞書、203は言語解析部、204は韻律生成規則保持部、205は韻律生成部、206は素片辞書、207は音声素片選択部、208は音声素片編集・接続部、209は音声波形出力部、210は音声データベース、211は素片辞書作成部、212はテキストコーパスである。このテキストコーパス212には、入力部106などを介して種々の内容のテキストを入力することができる。

#### 【0014】

まず、音声合成モジュールについて説明する。この音声合成モジュールでは、

言語解析部 203 が、解析辞書 202 を参照して、テキスト入力部 201 から入力されるテキストの言語解析を行なう。こうして解析された結果が韻律生成部 205 に入力される。韻律生成部 205 は、言語解析部 203 における解析結果と、韻律生成規則保持部 204 に保持されている韻律生成規則に関する情報とを基に、音韻系列と韻律情報を生成して音声素片選択部 207 及び音声素片編集・接続部 208 に出力する。続いて、音声素片選択部 207 は、韻律生成部 205 から入力される韻律生成結果を用いて、素片辞書 206 に保持されている音声素片から対応する音声素片を選択する。音声素片編集・接続部 208 は、韻律生成部 205 から入力される韻律生成結果に従って、音声素片選択部 207 から出力される音声素片を編集及び接続して音声波形を生成する。こうして生成された音声波形は、音声波形出力部 209 で出力される。

## 【0015】

次に、素片辞書作成モジュールについて説明する。

## 【0016】

このモジュールでは、素片辞書作成部 211 が、後述する手順に基づいて、音声データベース 210 の中から音声素片を選び出して素片辞書 206 に登録する。

## 【0017】

次に、上記構成を備えた本実施の形態の音声合成処理について説明する。

## 【0018】

図 3 は、図 2 の音声合成モジュールにおける音声合成処理（オンライン処理）の流れを示すフローチャートである。

## 【0019】

まずステップ S301 で、テキスト入力部 201 は、文、文節、単語等の単位毎にテキストデータを入力してステップ S302 に移る。ステップ S302 では、言語解析部 203 により当該テキストデータの言語解析を行う。次にステップ S303 に進み、音韻生成部 205 はステップ S302 で解析された結果と所定の韻律規則とに基づいて、音韻系列と韻律情報を生成する。次にステップ S304 に進み、各音韻毎にステップ S303 で得られた韻律情報と所定の音韻環境と

に基づいて、音声素片選択部207が素片辞書206に登録されている音声素片を選択する。次にステップS305に進み、その選択された音声素片及びステップS303で生成された韻律情報とに基づいて、音声素片編集・接続部208により音声素片の編集および接続を行なってステップS306に進む。ステップS306では、音声素片編集・接続部208によって生成された音声波形を、音声波形出力部209が音声信号として出力する。このようにして、入力されたテキストに対応する音声出力されることになる。

#### 【0020】

図4は、図2で示した素片辞書作成モジュールの、より詳細な構成を示すブロック図で、前述の図2と共通する部分は同じ番号で示し、かつ本実施の形態の特徴である素片辞書作成部211の構成をより詳細に示している。

#### 【0021】

図4において、401はテキスト入力部、402は言語解析部、403は解析辞書、404は韻律生成規則保持部、405は韻律生成部、406は音声素片検索部、407は音声素片保持部、408は音声素片編集部、409は変形歪決定部、410は接続歪決定部、411は歪決定部、412は歪保持部、413はNbest決定部、414はNbest保持部、415は登録素片決定部、416は登録素片保持部である。

#### 【0022】

以下、詳しく説明する。

#### 【0023】

テキスト入力部401は、テキストコーパス212から、例えば文単位にテキストデータを取り出して言語解析部402に出力する。言語解析部402は、解析辞書403を参照してテキスト入力部401から入力されたテキストデータを解析する。韻律生成部405は、言語解析部402で解析された解析結果に基づいて音韻系列を生成し、韻律生成規則保持部404が保持する韻律生成規則（アクセントパターン、自然降下成分、ピッチパターン等）を参照して韻律情報を生成する。音声素片検索部406は、韻律生成部405で生成される韻律情報と音韻系列とに従って音声データベース210から、各音韻毎に、所定の音韻環境を

考慮した音声素片を検索する。こうして検索された音声素片は一旦、音声素片保持部407に保持される。音声素片編集部408は、韻律生成部405で生成された韻律情報に合わせて音声素片保持部407に保持されている音声素片を編集する。この編集には、韻律情報に合わせて音声素片同士を接続する処理や、またその音声素片同士の接続に際して音声素片の一部を削除する等して変形する処理などが含まれる。

#### 【0024】

変形歪決定部409は、各音声素片の変形前と変形後の音響的特徴の変化から変形歪を決定する。接続歪決定部410は、音韻系列において一つ前の音声素片の終端付近の音響的特徴と当該音声素片の始端付近の音響的特徴から、これら音声素片同士が接続された場合の接続歪を決定する。歪決定部411は、変形歪決定部409で決定された変形歪と、接続歪決定部410で決定された接続歪とを考慮し、音韻系列ごとにトータルの歪（歪値ともいう）を決定する。歪保持部412は、歪決定部411で決定された各音声素片に至る歪の値を保持する。Nbest決定部413は、A\*（エースター）探索アルゴリズムを用いて、音韻系列毎に歪が最小となる上位N通りの最適パスを求める。Nbest保持部414は、Nbest決定部413で求めたN個の最適パスを入力テキストごとに保持する。登録素片決定部415は、Nbest保持部414に保持されている、各音韻ごとにNbestの結果から、その頻度順に、素片辞書206に登録する音声素片を選び出す。登録素片保持部416は、登録素片決定部415により選ばれた音声素片を保持する。

#### 【0025】

図5は、図4で示す素片辞書作成モジュールにおける処理の流れを示すフローチャートである。

#### 【0026】

まずステップS501で、テキスト入力部401がテキストコーパス212から一文ずつテキストデータを取り出す。取り出せるテキストデータが存在しなくなると、最終的に登録する音声素片を決定するステップS512に進む。テキストデータが存在する場合はステップS502に進み、言語解析部402において

、解析辞書403を使って、その入力されたテキストデータの言語解析を行なってステップS503に進む。ステップS503では、韻律生成部405により、韻律生成規則保持部404が保持する韻律生成規則と、ステップS502における言語解析結果とに基づいて韻律情報並びに音韻系列を生成する。次にステップS504に進み、ステップS503で生成された音韻系列内の各音韻を順次処理する。このステップS504で未処理の音韻が存在しない場合はステップS511に進むが、未処理の音韻が存在する場合はステップS505に進む。ステップS505において、音声素片検索部406は、各音韻毎に音韻環境及び韻律規則を満足する音声素片を音声データベース210から検索して音声素片保持部407に保存する。

## 【0027】

例えば具体例で説明すると、テキストデータとして「こんにちわ」が入力されると、それが言語解析され、アクセントやイントネーション等を含む韻律情報が生成される。そして、この「こんにちわ」は、例えばd i p h o n eを音韻の単位として用いた場合、以下のような音韻系列に分解される。

## 【0028】

こ            ん            に            ち            わ

/k k.o o.X X.n n.i i.t t.i i.w w.a a/

なお、ここで「X」は、音声「ん」を示し、「/」は無声音を示す。

## 【0029】

次にステップS506に進み、その検索された複数の音声素片について順次処理する。未処理の音声素片が存在しない場合はステップS504に戻って次の音韻の処理に進むが、存在する場合はステップS507に進んで、現在の音韻の音声素片を処理する。ステップS507では、音声素片編集部408が、上述の音声合成処理時と同じ手法を用いて音声素片の編集を行なう。ここでいう音声素片の編集とは、例えばピッチ同期波形重畳法（PSOLA）などの処理である。この音声素片の編集には、その音声素片と韻律情報を用いる。音声素片の編集が終了したらステップS508に進み、変形歪決定部409により、現在の音声素片の変形前と変形後における音響的特徴の変化を変形歪として算出する（この詳細は後



述する)。次にステップ S 5 0 9 に進み、接続歪決定部 4 1 0 により、現在の音声素片とその一つ前の音韻の音声素片の全てとの接続歪を算出する(この処理についても詳しく後述する)。次にステップ S 5 1 0 に進み、歪決定部 4 1 1 は、変形歪と接続歪から現在の音声素片に至るパスの全てについて歪値を決定する(後述する)。そして現在の音声素片に至るパスの歪値の上位 N 個 (N : 求めたい Nbest の個数) と、そのパスを表わす一つ前の音韻の音声素片へのポインタを歪保持部 4 1 2 に保持してステップ S 5 0 6 に戻り、現在の音韻において未処理の音声素片が存在するかどうかを調べる。

## 【 0 0 3 0 】

こうしてステップ S 5 0 6 で、各音韻における全ての音声素片が処理され、更にステップ S 5 0 4 で全ての音韻が処理されるとステップ S 5 1 1 に進む。ステップ S 5 1 1 において、Nbest 決定部 4 1 3 は、A\*探索アルゴリズムを用いた Nbest 探索を行ない、上位 N 位までの最適パス(音声素片系列ともいう)を求め、これを Nbest 保持部 4 1 4 に保持してステップ S 5 0 1 に戻る。

## 【 0 0 3 1 】

こうして全テキストに対する処理が終了するとステップ S 5 0 1 からステップ S 5 1 2 に進み、登録素片決定部 4 1 5 は、音韻ごとに全テキストの Nbest 結果に基づいて所定の頻度の高い以上を選択して音声素片を素片辞書 2 0 6 に登録する。尚、この Nbest における N の値は、予備実験などから経験的に与えておく。こうして決定された音声素片は、登録素片保持部 4 1 6 を介して素片辞書 2 0 6 に登録される。

## 【 0 0 3 2 】

図 6 は、本実施の形態に係る図 5 のステップ S 5 0 8 における変形歪の求め方を説明する図である。

## 【 0 0 3 3 】

ここでは、PSOLA 法によりピッチ間隔を広げる場合について図示している。矢印はピッチマーク、点線は変形前と変形後のピッチ素片の対応関係を表わしている。本実施の形態では、各ピッチ素片(微細素片ともいう)の変形前後のケプストラム距離に基づいて変形歪を表わす。具体的には、まず変形後のあるピッチ素

片（例えば60で示す）のピッチマーク61を中心にハニング窓62（窓長25.6ミリ秒）をかけ、そのピッチ素片60を周辺のピッチ素片を含めて切り出す。こうして切り出したピッチ素片60をケプストラム分析する。次に、ピッチマーク61に対応する変形前のピッチ素片63のピッチマーク64を中心にして同じ窓長のハニング窓65でピッチ素片を切り出し、変形後の場合と同様にしてケプストラムを求める。このようにして求めたケプストラム同士の距離を、着目しているピッチ素片60の変形歪として、変形後のピッチ素片とそれに対応する変形前のピッチ素片間の変形歪の総和をPSOLAで採用されるピッチ素片数 $N_p$ で割った値を、その音声素片の変形歪とする。こうして求められる変形歪を式で記述すると以下ようになる。

【0034】

$$Dt = \sum \sum |Corg\ i,j - Ctar\ i,j| / Np$$

ここで最初の $\sum$ は、 $i = 1$ から $N$ までの総和を示し、次の $\sum$ は $j = 0 \sim 16$ までの総和を示している。また $Ctar\ i,j$ は、変形後の $i$ 番目のピッチ素片のケプストラムの $j$ 次元目の要素を表わし、同様に、 $Corg\ i,j$ は、変形後に対応する変形前のピッチ素片のケプストラムの $j$ 次元目の要素を表わしている。

【0035】

図7は、本実施の形態における接続歪の求め方を説明する図である。

【0036】

この接続歪は、一つ前の音韻の音声素片と現在の音声素片との接続箇所において生じる歪を示し、ここではケプストラム距離を用いて表わす。具体的には、音声素片境界が存在するフレーム70, 71（フレーム長5ミリ秒、分析窓幅25.6ミリ秒）と、それを挟む前後それぞれ2フレームからなる計5フレームを接続歪の算出対象としている。ここでケプストラムは、0次（パワー）～16次（パワー）までの計17次元ベクトルとする。そして、このケプストラムベクトルの各要素の差の絶対値の和を、現在注目している音声素片における接続歪とする。即ち、図7の700で示すように、一つ前の音韻の音声素片における終端部のケプストラムベクトルの各要素を $Cpre\ i,j$ （ $i$ ：フレーム番号、フレーム番号の“0”が音声素片境界があるフレームを示し、 $j$ がベクトルの要素番号を示す

）とする。また、図 7 の 7 0 1 で示すように、注目音声素片における始端部のケプストラムベクトルの各要素を  $C_{cur\ i,j}$  とすると、現在注目している音声素片の接続歪  $D_c$  は、

$$D_c = \sum_i \sum_j |C_{pre\ i,j} - C_{cur\ i,j}|$$

となる。ここで最初の  $\sum$  は  $i = -2 \sim 2$  の総和を、次の  $\sum$  は  $j = 0 \sim 16$  までの総和を示す。

#### 【0037】

図 8 は、本実施の形態に係る歪決定部 4 1 1 による、音声素片における歪の決定過程を図示したものである。本実施の形態において、音韻単位は *diphone*（ダイフォン）とする。

#### 【0038】

図中、一つの円がある音韻における 1 つの音声素片を示し、円内の数字は、この音声素片に至る歪値の総和の最小値を示している。また四角で囲まれた数字は、一つ前の音韻の音声素片と現在注目している音韻の音声素片との間の歪値を示している。また矢印は、現在注目している音韻の音声素片と一つ前の音韻の音声素片との関連を示している。ここでは説明のため、 $n$  番目の音韻（現在注目している音韻）の  $m$  番目の音声素片を  $P_{n,m}$  とする。この音声素片  $P_{n,m}$  の最も小さい歪値から上位  $N$  個（ $N$ ：求めたい  $N_{best}$  の数）までに対応する音声素片を一つ前の音韻の中から取り出し、その中の  $k$  番目の歪値を  $D_{n,m,k}$  とし、その歪値に対応するの一つ前の音韻の音声素片を  $PRE_{n,m,k}$  とすると、 $PRE_{n,m,k}$  を介して音声素片  $P_{n,m}$  に至るパスにおける歪値の総和  $S_{n,m,k}$  は、

$$S_{n,m,k} = S_{n-1,x,0} + D_{n,m,k} \quad (\text{但し、} x = PRE_{n,m,k})$$

となる。

#### 【0039】

本実施の形態における歪値について説明する。本実施の形態では歪値  $D_{total}$ （上記説明における  $D_{n,m,k}$  に相当する）を、上述の接続歪  $D_c$  と変形歪  $D_t$  の重み付き和として定義する。

#### 【0040】

$$D_{total} = w \times D_c + (1 - w) \times D_t \quad : (0 \leq w \leq 1)$$

ここで重み係数  $w$  は、予備実験など経験的に求められる係数で、 $w = 0$  の場合は、歪値が変形歪  $D_t$  のみで説明され、 $w = 1$  の場合は、歪値が接続歪  $D_c$  のみに依存することになる。

#### 【0041】

歪保持部 412 では、各音韻の音声素片  $P_{n,m}$  毎に、上位  $N$  個の歪値  $D_{n,m,k}$  と、それらに対応する一つ前の音韻の音声素片  $PRE_{n,m,k}$  と、 $PRE_{n,m,k}$  を介して  $D_{n,m,k}$  に至るパスの歪値の総和  $S_{n,m,k}$  をそれぞれ保持する。

#### 【0042】

図 8 では、現在注目している音声素片  $P_{n,m}$  に至るパスの総和の最小値が「222」となる例を示す。この時の音声素片  $P_{n,m}$  の歪値は、 $D_{n,m,1} (k=1)$  であり、この歪値  $D_{n,m,1}$  に対応する一つ前の音韻の音声素片は、 $PRE_{n,m,1}$  (図 8 の  $P_{n-1,m,8,1}$  に相当する) である。80 は、音声素片  $PRE_{n,m,1}$  と音声素片  $P_{n,m}$  とを接続するパスである。

#### 【0043】

図 9 は、 $N_{best}$  の決定過程を図示したものである。

#### 【0044】

ステップ S510 の終了時点で、各音声素片において、上位  $N$  個の情報がそれぞれ求まっている (フォワード探索)。  $N_{best}$  決定部 413 では、音韻系列の末尾の音声素片 90 から逆順に枝を伸ばしながら  $N_{best}$  パスを求める (バックワード探索)。この枝を伸ばすノードの選択は、予測値 (線の横の数字) とそこに至る総歪値の和 (歪値は四角の中の数字で示される) が最小となるものである。ここでいう予測値とは、音声素片  $P_{n,m}$  におけるフォワード探索結果の最小歪  $S_{n,m,0}$  に相当する。この場合、予測値と実際に左端までに至る最小パスの歪が等しいので、A\*探索アルゴリズムの性質により最適パスが求まることが保証される。

#### 【0045】

図 9 は、第 1 位の最適パスが決定された状態を示す図である。

#### 【0046】

図中、丸が音声素片を示し、その丸の中の数字が歪み予測値、太い実線が第一位のパス、四角の中の数字が歪値、線の横の数字が予測歪み値を示している。次

に第2位のパスを求めるために、二重丸のノードの中で、予測値とそこに至る総歪値の和が最小となるノードを選択し、それに繋がる一つ前の音韻の音声素片の全て（最大N個）に枝を伸ばす。この伸ばした先のノードが二重丸で表現されている。この操作を繰り返すことにより、上位N個のパスが総歪値の順に決定される。この図9は、N=2として枝を伸ばした場合の例を示す図である。

【0047】

このようにして本実施の形態1によれば、歪の最も小さいパスを形成する音声素片を選択して、それを素片辞書に登録することができる。

【0048】

〔実施の形態2〕

前述の実施の形態1では、音韻の単位としてdiphoneを用いる場合について記述したが、本発明はこれに限定されるものではなく、音素や半diphoneなどを単位としてもよい。半diphoneとは、diphoneを音素境界で2つに分割したものである。この半diphoneを単位とした場合のメリットについて簡単に説明する。任意のテキストを合成する場合、素片辞書206は全種類のdiphoneを用意しておく必要がある。これに対して、半diphoneを単位とした場合は、足りない半diphoneを別の半diphoneで代替できる。例えば、半diphoneの「/a.b.0/(diphone a.bの左側)」の代わりに「/a.n.0/」を利用しても、音質の劣化を少なくして良好に音声を再生できる。これにより、素片辞書206のサイズをより小さくできる。

【0049】

〔実施の形態3〕

前述の実施の形態1、2では、音韻の単位としてdiphoneや音素や半diphoneを用いる場合について説明したが、本発明はこれに限定されるものではなく、これらを混合して用いてもよい。例えば、利用頻度が高い音韻については、diphoneを単位とし、利用頻度が低い音韻については、2つの半diphoneを用いて表現するようにしても良い。

【0050】

図10は、音声素片単位を混合した場合の一例を示した図で、ここでは音韻「

o.w] がdiphoneで表され、その前後の音韻は半diphoneで表されている。

【0051】

〔実施の形態4〕

実施の形態3において、元のデータベース中で連続する場所から取り出されたかどうかの情報を持ち、連続していた場合は、半diphoneの組を仮想的にdiphoneとして扱うようにしてもよい。つまり、元のデータベース中で連続するということは接続歪が“0”であるため、この場合には変形歪だけを考慮すればよいことになり計算量を大幅に軽減できる。

【0052】

図11は、この様子を表わした概念図である。図中の線上の数字は接続歪を表している。

【0053】

図11において、1100で示される半diphoneの組は、元のデータベース中で連続する場所から取り出されたものであり、その接続歪みは“0”に一義的に決定されている。また1101で示された半diphoneの組は、元のデータベース中で連続する場所から取り出されたものではないため、それぞれに対して接続歪みが計算される。

【0054】

〔実施の形態5〕

前述の実施の形態1では、1単位のテキストデータから得られた音韻系列全体を歪計算の対象とする場合について説明したが、本発明はこれに限定されるものでない。例えば、ポーズや無音部分までを一つの区間として音韻系列を分割し、各区間ごとに歪計算を行ってもよい。ここで言う無音部分とは、例えばp,t,kなどの無音部分のことである。ポーズや無音部分では接続歪が“0”であると考えられるため、このような分割が有効となる。これにより、各区間毎に最適な音声素片の選択が可能となる。

【0055】

〔実施の形態6〕

前述の実施の形態1では、接続歪の計算にケプストラムを用いる場合について

説明したが、本発明はこれに限定されるものではない。例えば、接続点の前後に互る波形の差分の和を用いて接続歪を求めても良い。またスペクトル距離などを用いて接続歪を求めてもよい。この場合、接続点はピッチマークに同期させるのが、より好ましい。

【0056】

〔実施の形態7〕

前述の実施の形態1では、接続歪の計算において、窓長、シフト長、ケプストラムの次数、フレーム数などを具体的数字を使って説明したが、本発明はこれに限定されるものではない。任意の窓長、シフト長、次数、フレーム数を使って接続歪を算出してもよい。

【0057】

〔実施の形態8〕

前述の実施の形態1では、接続歪の計算にケプストラムの次数ごとに差分を取ったものの総和を用いる場合について説明したが、本発明はこれに限定されるものではない。例えば、各次数を統計的性質などを使って正規化（正規化係数  $r_j$ ）してもよい。この場合の接続歪  $D_c$ は、

$$D_c = \sum \sum (r_j \times |C_{pre\ i,j} - C_{cur\ i,j}|)$$

となる。ここで、最初の  $\sum$  は  $i = -2 \sim 2$  の総和を、次の  $\sum$  は  $j = 0 \sim 16$  までの総和を示す。

【0058】

〔実施の形態9〕

実施の形態1では、ケプストラムの次数ごとの差分の絶対値をベースに接続歪の算出を行なう場合について説明したが、本発明はこれに限定されるものではない。例えば、差分の絶対値の累乗（累数が偶数の場合は絶対値でなくてもよい）をベースに接続歪の算出を行なってもよい。ここで累数を  $N$  とすると、接続歪  $D_c$  は、

$$D_c = \sum \sum |C_{pre\ i,j} - C_{cur\ i,j}|^N$$

となる。ここで“ $N$ ”は  $N$  の累乗を示す。ここで  $N$  の値を大きくすることは、大きな差分について敏感になることを意味しているので、その結果、接続歪が平均

的に小さくなるように働くことになる。

#### 【0059】

##### 〔実施の形態10〕

前述の実施の形態1では、変形歪としてケプストラムを用いる場合について説明したが、本発明はこれに限定されるものではない。例えば、変形前後の一定区間の波形の差分の和を用いて変形歪を求めてもよい。また、スペクトル距離などを用いて変形歪を求めてもよい。

#### 【0060】

##### 〔実施の形態11〕

前述の実施の形態1では、変形歪を波形から得られる情報を基に算出する場合について説明したが、本発明はこれに限定されるものではない。例えば、PSOLAによるピッチ素片の削除および複製の回数などを変形歪を算出する要素としても良い。

#### 【0061】

##### 〔実施の形態12〕

前述の実施の形態1では、音声素片を読み出すごとに接続歪を計算する場合について説明したが、本発明はこれに限定されるものではない。例えば、接続歪を予め計算しておき、テーブル化して保持してもよいものとする。

#### 【0062】

図12は、diphone「/a.r/」とdiphone「/r.i/」との間の接続歪を記憶したテーブルの一例を示す図である。ここでは縦軸に「/a.r/」の音声素片、横軸に「/r.i/」の音声素片をとっている。例えば、「/a.r/」の「id3」の音声素片と「/r.i/」の「id2」の音声素片との接続歪は“3.6”で表されている。このように接続可能なdiphone間の接続歪を全てテーブル化して用意することにより、音声素片同士の合成時の接続歪の算出がテーブルの参照だけで済むため、その計算量を大幅に軽減でき、算出時間を大幅に短縮できる。

#### 【0063】

##### 〔実施の形態13〕

前述の実施の形態1では、音声素片を編集する毎に変形歪を計算する場合につ



いて説明したが、本発明はこれに限定されるものではない。例えば、変形歪を予め計算しておき、テーブルとして保持しておいても良い。

## 【 0 0 6 4 】

図 1 3 は、あるdiphoneを基本周波数と音韻時間長について変化させた場合の変形歪をテーブルで表した図である。

## 【 0 0 6 5 】

図中、 $\mu$  は、そのdiphoneの統計的な平均値を示し、 $\sigma$  は標準偏差である。具体的な表の作成方法としては、次のような作成方法が考えられる。まず、基本周波数と音韻時間長に関して統計的に平均値と分散を求める。次に、それらを基に（ $5 \times 5 =$ ）25通りの基本周波数と音韻時間長をターゲットとしてPSOLA法をそれぞれ適用し、テーブルの変形歪を一つずつ求めていけばよい。合成時は、ターゲットの基本周波数と音韻時間長が決まれば、テーブルの近傍の値で内挿（もしくは外挿）することによって、変形歪を推定することが可能である。

## 【 0 0 6 6 】

図 1 4 は、合成時に変形歪を推定するための具体例を示した図である。

## 【 0 0 6 7 】

図中、黒丸がターゲットの基本周波数と音韻時間長であり、このとき、各格子点の変形歪がテーブルからA, B, C, Dと求まっていると仮定すると、変形歪Dtは、以下の式により求めることができる。

$$Dt = \{A \cdot (1 - y) + C \cdot y\} \times (1 - x) + \{B \cdot (1 - y) + D \cdot y\} \times x$$

## 【 0 0 6 8 】

## 〔実施の形態 1 4〕

前述の実施の形態 1 3 では、変形歪テーブルの格子点として、そのdiphoneの統計的な平均値と標準偏差を基に  $5 \times 5$  のテーブルを作成したが、本発明はこれに限定されるものではなく、任意の格子点を持つテーブルとしてもよい。また、格子点を平均値などに依らず決定的に与えてもよいものとする。例えば、韻律推定で推定されうる範囲を等分割するなどもよいものとする。

## 【 0 0 6 9 】

## 〔実施の形態 1 5〕

前述の実施の形態1では、接続歪と変形歪の重み和で歪を定量化する場合について説明したが本発明はこれに限定されるものではなく、接続歪と変形歪それぞれに閾値を設定しておき、どちらか一方でもその閾値を越えた場合はその音声素片が選択されないようにして、十分大きな歪の値を与えるようにしてもよい。

【0070】

上記実施の形態においては、各部を同一の計算機上で構成する場合について説明したが本発明はこれに限定されるものではなく、例えばネットワーク上に分散した計算機や処理装置などに分かれて各部を構成してもよい。

【0071】

上記実施の形態においては、プログラムを制御メモリ（ROM）に保持する場合について説明したが本発明はこれに限定されるものではなく、外部記憶など任意の記憶媒体を用いて実現してもよい。また、同様の動作をする回路で実現してもよい。

【0072】

なお本発明は、複数の機器から構成されるシステムに適用しても、1つの機器からなる装置に適用してもよい。前述した実施の形態の機能を実現するソフトウェアのプログラムコードを記録した記録媒体を、システム或いは装置に供給し、そのシステム或いは装置のコンピュータ（またはCPUやMPU）が記録媒体に格納されたプログラムコードを読み出し実行することによっても達成される。

【0073】

この場合、記録媒体から読み出されたプログラムコード自体が前述した実施の形態の機能を実現することになり、そのプログラムコードを記録した記録媒体は本発明を構成することになる。プログラムコードを供給するための記録媒体としては、例えば、フロッピーディスク、ハードディスク、光ディスク、光磁気ディスク、CD-ROM、CD-R、磁気テープ、不揮発性のメモリカード、ROMなどを用いることができる。

【0074】

また、コンピュータが読み出したプログラムコードを実行することにより、前述した実施の形態の機能が実現されるだけでなく、そのプログラムコードの指示

に基づき、コンピュータ上で稼働しているOSなどが実際の処理の一部または全部を行ない、その処理によって前述した実施の形態の機能が実現される場合も含まれる。

#### 【0075】

更に、記録媒体から読み出されたプログラムコードが、コンピュータに挿入された機能拡張ボードやコンピュータに接続された機能拡張ユニットに備わるメモリに書き込まれた後、そのプログラムコードの指示に基づき、その機能拡張ボードや機能拡張ユニットに備わるCPUなどが実際の処理の一部または全部を行ない、その処理によって前述した実施の形態の機能が実現される場合も含まれるものとする。

#### 【0076】

以上説明したように本実施の形態によれば、接続歪と変形歪を考慮して素片辞書に登録する音声素片を選択することにより、少数の音声素片に登録した辞書を用いても、音質の劣化が少ない合成音声を生成できるという効果がある。

#### 【0077】

##### 【発明の効果】

以上説明したように本発明によれば、接続歪や変形歪に基づく歪の影響を考慮して素片辞書に登録する音声素片を選択することによって、そのような素片辞書を用いた合成音声の質を向上できるという効果がある。

#### 【0078】

また本発明によれば、素片辞書に登録する音声素片の数を少なく抑えて、かつその素片辞書を用いて良好な音声を再生できるという効果がある。

##### 【図面の簡単な説明】

##### 【図1】

本発明の実施の形態に係る音声情報処理装置のハードウェア構成を示すブロック図である。

##### 【図2】

本発明の実施の形態1に係る音声情報処理装置のモジュール構成を示すブロック図である。

【図 3】

本実施の形態に係るオンラインモジュールにおける処理の流れを示すフローチャートである。

【図 4】

本実施の形態に係るオフラインモジュールの詳細な構成を示すブロック図である。

【図 5】

本実施の形態 1 に係るオフラインモジュールにおける処理の流れを示すフローチャートである。

【図 6】

本発明の実施の形態に係る音声素片の変形を説明する図である。

【図 7】

本発明の実施の形態に係る音声素片の接続歪を説明する図である。

【図 8】

音声素片における歪の決定過程を説明する図である。

【図 9】

Nbestによる決定過程を説明する図である。

【図 10】

本発明の実施の形態 3 に係る音声素片の単位をdiphoneと半diphoneとで混合した場合を説明する図である。

【図 11】

本発明の実施の形態 4 に係る音声素片の単位を取り出した半diphoneによって混合した例を示した図である。

【図 12】

本発明の実施の形態 12 に係るdiphoneの /a.r/ と /r.i/ 間の接続歪を決定するテーブル構成例を示す図である。

【図 13】

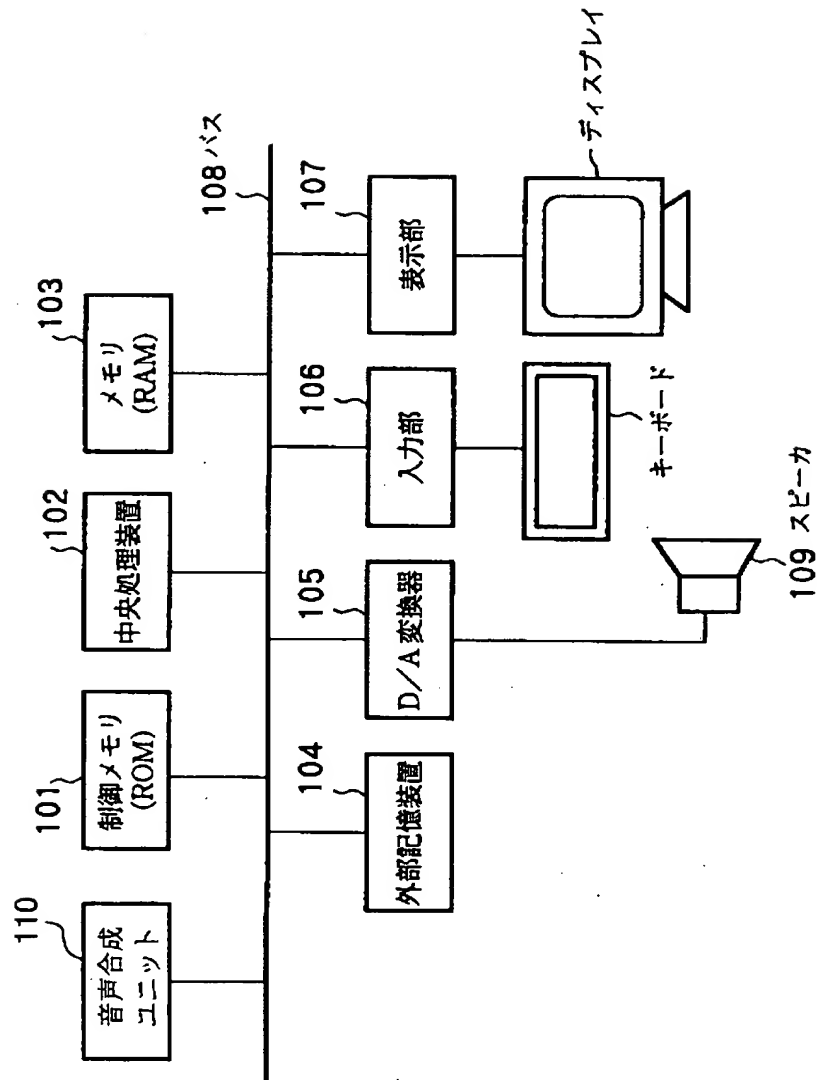
本発明の実施の形態 13 に係る変形歪を表わすテーブル例を示す図である。

【図 14】

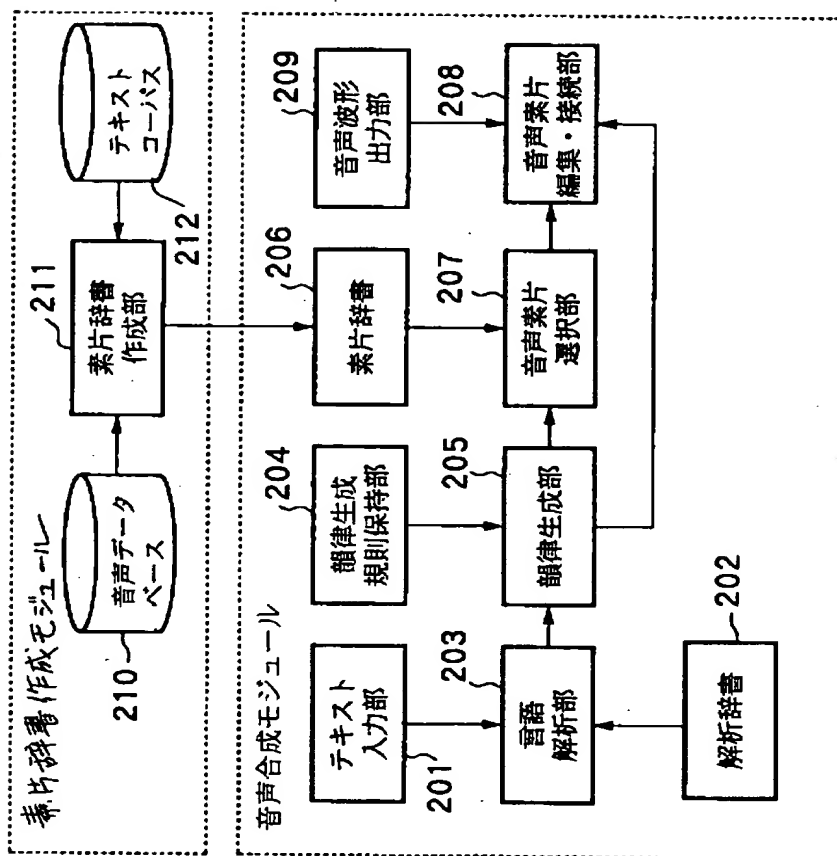
本発明の実施の形態 1 3 に係る変形歪を推定する具体例を示した図である。

【書類名】 図面

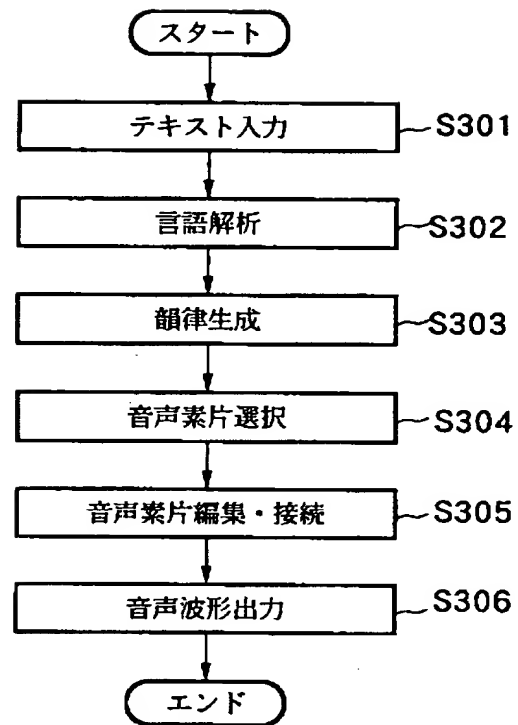
【図 1】



【図 2】

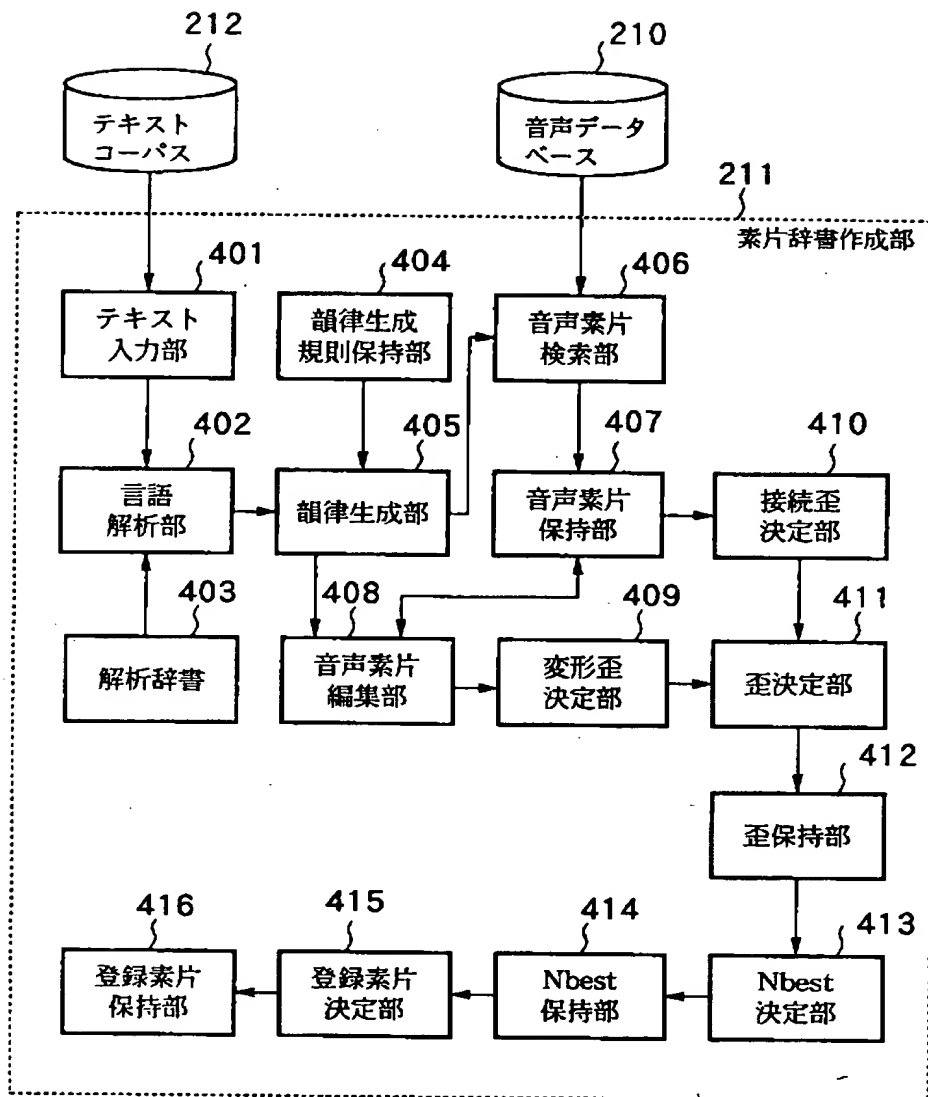


【図 3】

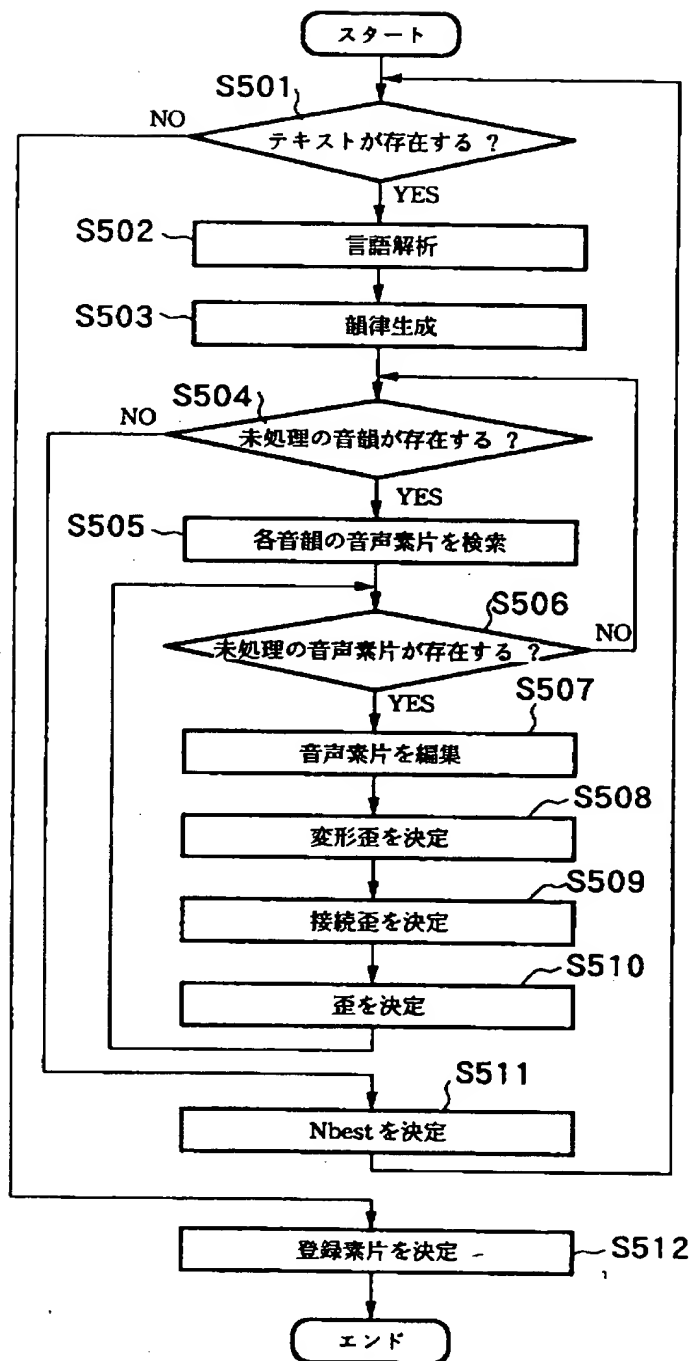




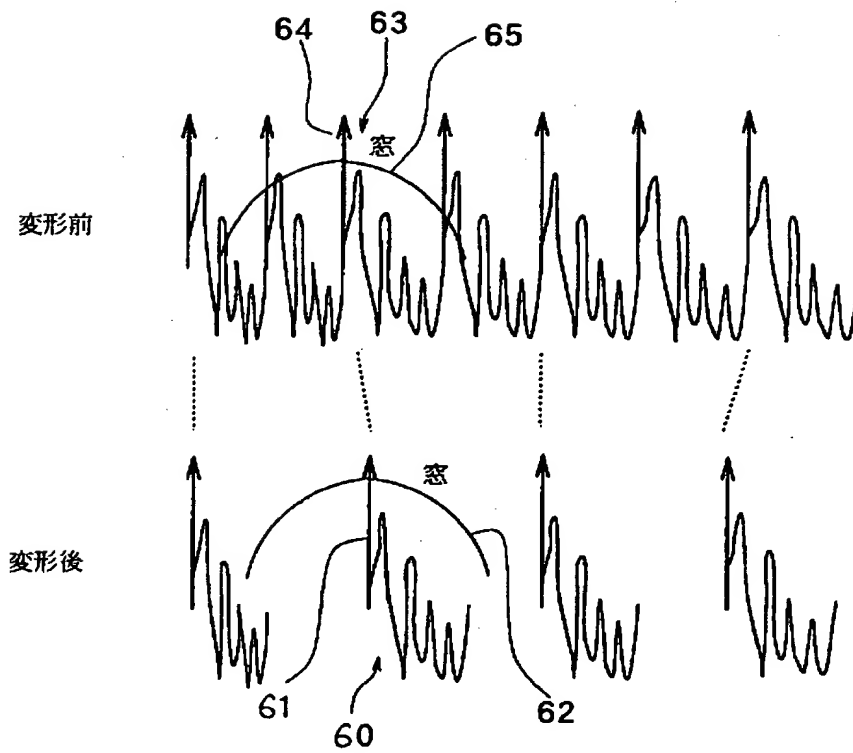
【図 4】



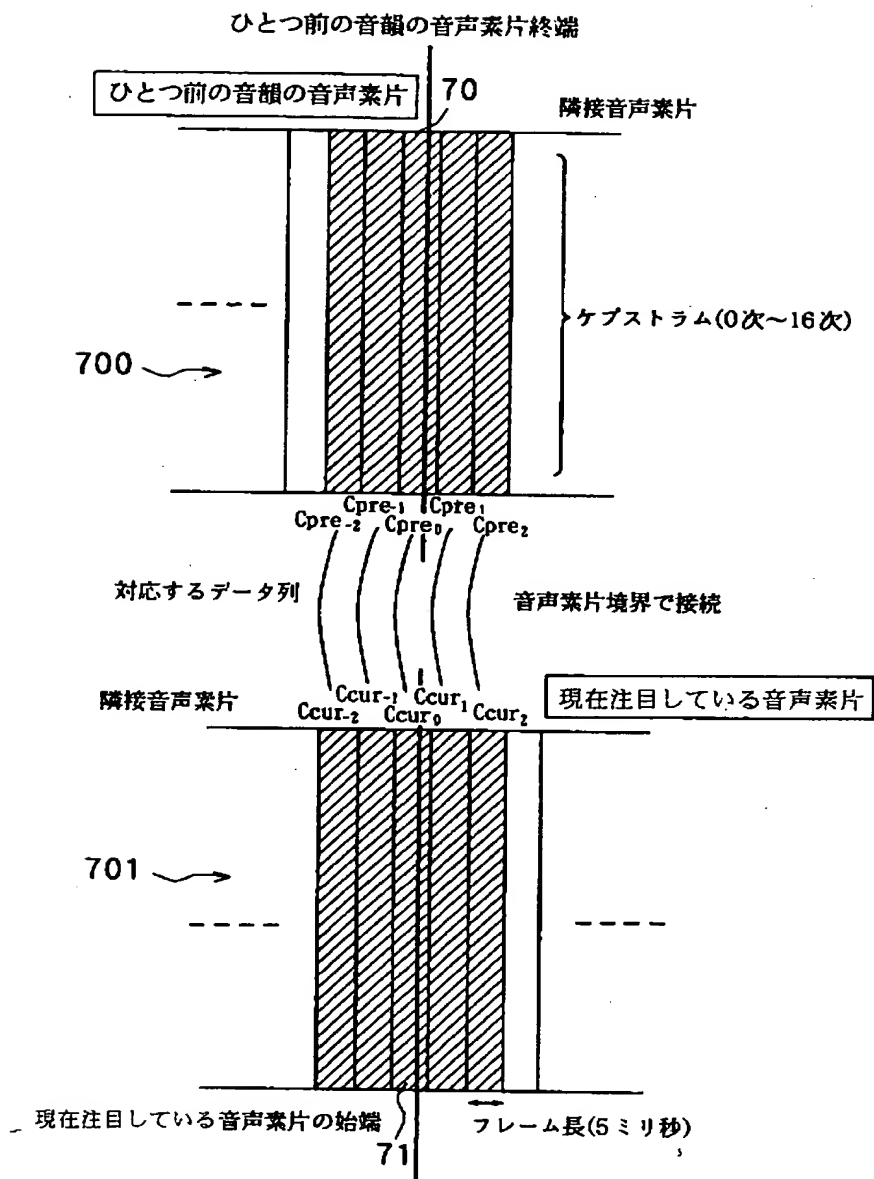
【図 5】



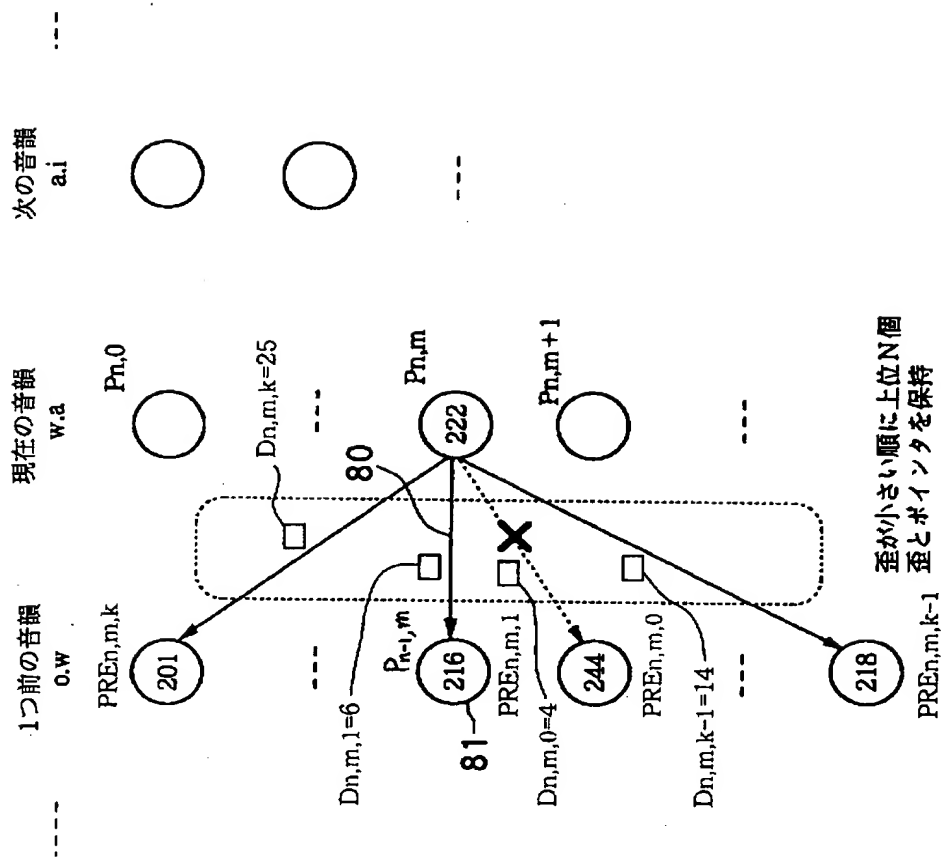
【図 6】



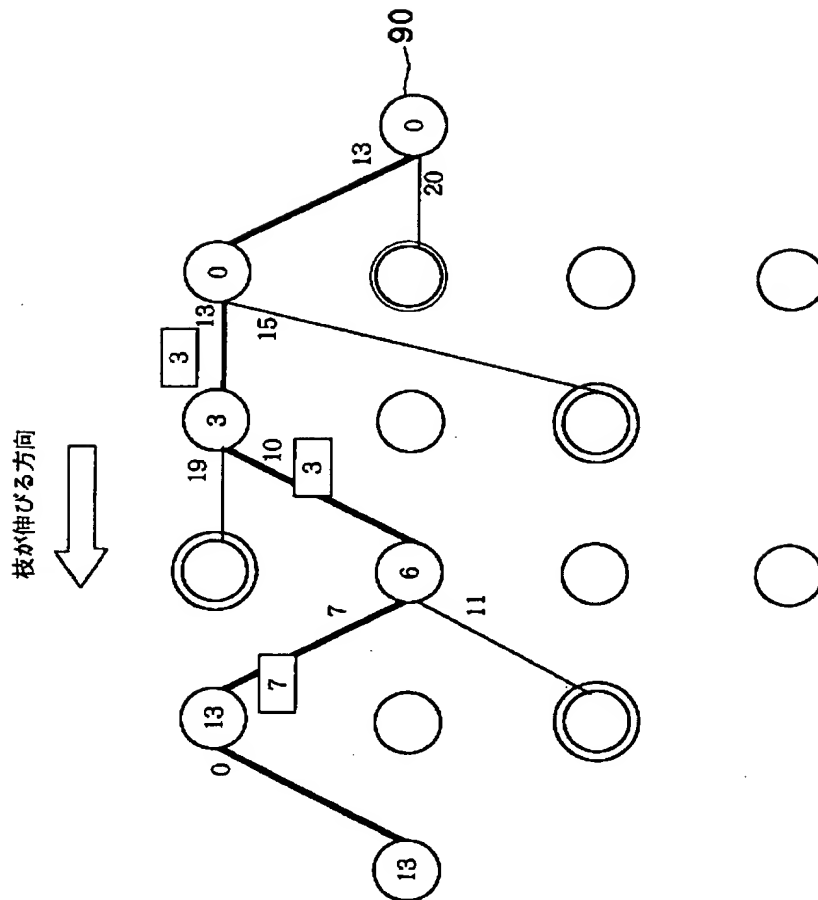
【図 7】



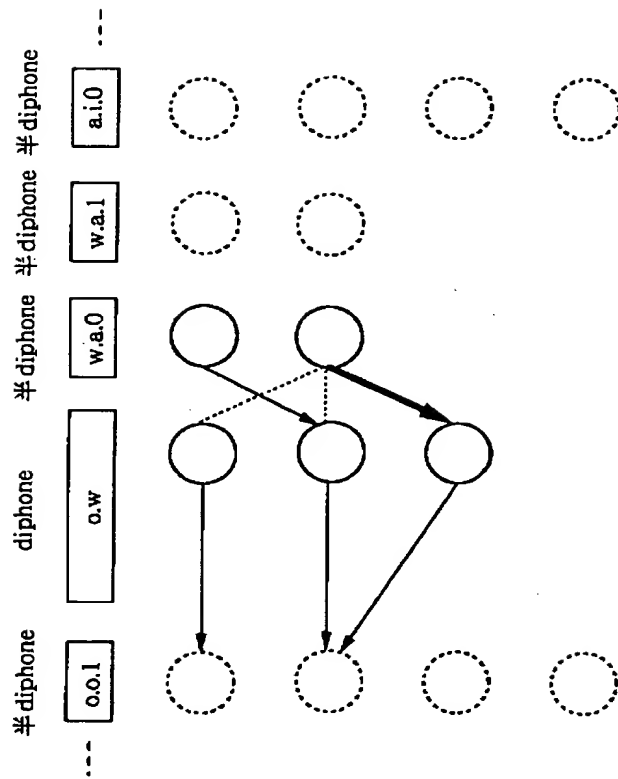
【図 8】



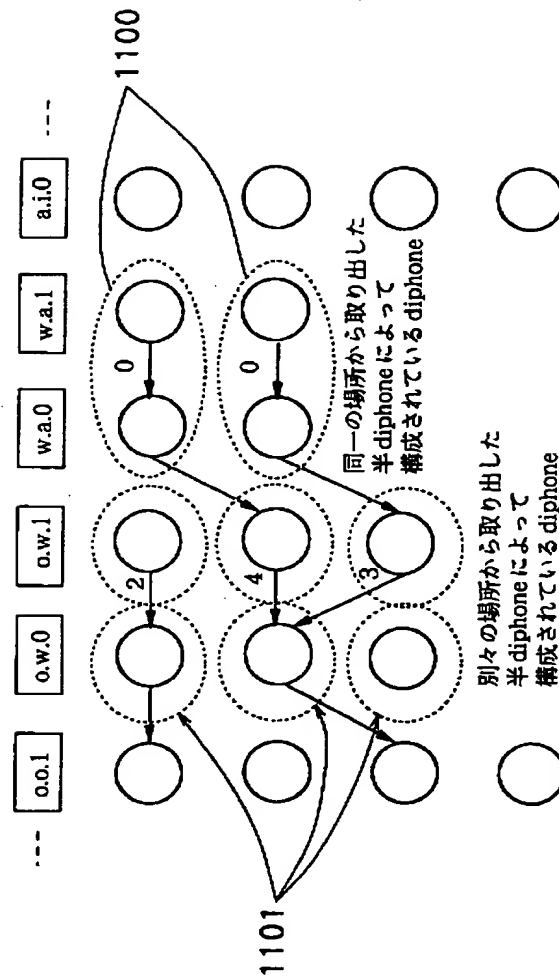
【図 9】



【図 10】



【図 1 1】





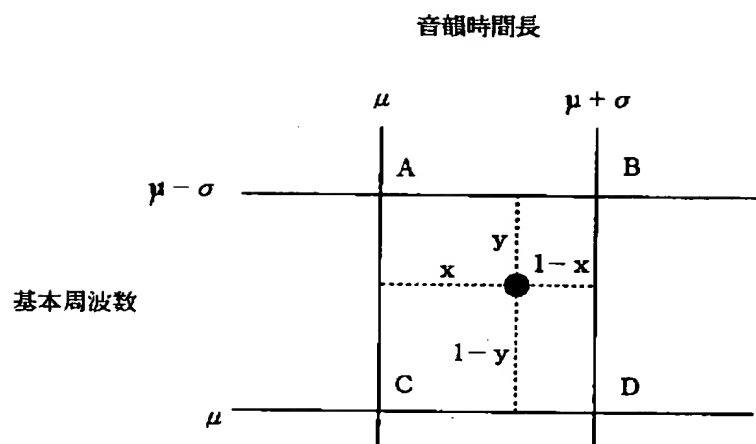
【図 1 2】

$\begin{matrix} r.i \\ a.r \end{matrix}$	id1	id2	id3	id4	id5	...
id1	13.2	2.6	...			
id2	22	11.2				
id3	5	3.6				
id4	0	4				
id5	3.3	...				
...						

【図 13】

		音韻時間長				
基本周波数		$\mu - 2\sigma$	$\mu - \sigma$	$\mu$	$\mu + \sigma$	$\mu + 2\sigma$
	$\mu - 2\sigma$	26	25	15	13	14
	$\mu - \sigma$	20	24	19	18	17
	$\mu$	15	19	10	13	20
	$\mu + \sigma$	19	22	16	22	26
	$\mu + 2\sigma$	25	27	26	30	31

【図 14】



【書類名】 要約書

【要約】

【課題】 素片辞書に登録する音声素片の数を少なく抑えて、かつその素片辞書を用いて良好な音声を再生する。

【解決手段】 入力したテキストデータを言語解析して韻律を生成し、その韻律に基づいて音声データベース 2 1 0 から音声素片を検索する。この検索された音声素片の変形歪、及び一つ前の音韻の音声素片との接続による接続歪を求め、歪決定部 4 1 1 により、変形歪と接続歪の重み付け等を行なってトータルの歪を決定する。次に Nbest 決定部 4 1 3 により、A\* (エースター) 探索アルゴリズムを用いて歪が最小となる上位 N 通りの最適パスを求め、登録素片決定部 4 1 5 は上位 N 通りの最適パスから、その頻度順に素片辞書 2 0 6 に登録する登録素片を選び出し、それを素片辞書に登録する。

【選択図】 図 4

出 願 人 履 歴 情 報

識別番号 [000001007]

1. 変更年月日	1990年 8月30日
[変更理由]	新規登録
住 所	東京都大田区下丸子3丁目30番2号
氏 名	キャノン株式会社